FP7 Project nr	226487	Project start date:	01 May 2009
Acronym:	EuroGEOSS		
Project title:	EuroGEOSS, a European approach to GEOSS		
Theme:	FP7-ENV-2008-1: Environment (including climate change)		
Theme title:	ENV.2008.4.1.1.1: European Environment Earth Observation system supporting INSPIRE and compatible with GEOSS (Global Earth Observation System of Systems)		
Web site:	www.eurogeoss.eu		

D.5.2b Update on metadata and catalogue: the Drought Vocabulary description and development process

Title	D.5.2b Update on metadata and catalogue: the Drought Vocabulary description and development process
Creator	Miguel Ángel Latre, Barbara Hofer, Javier Lacasta, Javier Nogueras-Iso
Creation date	11/11/2011
Date of last revision	15/11/2011
Subject	Metadata, annotation, vocabularies, appropriateness, drought
Status	☐ Draft ☒ Final
Publisher	EuroGEOSS
Type	Text
Description	Description of the Drought Vocabulary developed within the WP5, and its creation process
Contributor	Barbara Medved-Cvikl, Andrej Ceglar, Rogelio Galván, Carolina de Carvalho, Florian Husson, Jean-Francois Vernoux, José Miguel Rubio, Stefan Niemeyer, Jürgen Vogt
Format	PDF
Source	
Rights	☐ Restricted ☒ Public
Identifier	D.5.2b.doc
Language	En
Relation	WP5
Coverage	Not applicable

These are Dublin Core metadata elements. See for more details and examples http://www.dublincore.org/

TABLE OF CONTENTS

FIGURES

TABLES

ACRONYMS AND ABBREVIATIONS

Abbreviation	Name
AAT	Art & Architecture Thesaurus
AMNH	American Museum of Natural History
ANSI	American National Standards Institute
BSI	British Standards Institution
CUAHSI	Consortium of Universities for the Advancement of Hydrologic Science, Inc.
DIRKS	Designing and Implementing Recordkeeping Systems
EDO	European Drought Observatory
GEMET	General Multilingual Environmental Thesaurus
GEOSS	Global Earth Observation System of Systems
IJSDIR	International Journal of Spatial Data Infrastructures Research
INSPIRE	Infrastructure for Spatial Information in Europe
ISO	International Organization for Standardization
JRC	Joint Research Centre
JRC-IES RDSI	Reference Data and Services Initiative of the JRC Institute for Environment and Sustainability
NISO	National Information Standards Organization
OGC	Open Geospatial Consortium
RDF	Resource Description Format
RDFS	RDF Schema
SBA	GEOSS Societal Benefit Areas
SESAME	Project "Southern European Seas: Assessing and Modelling Ecosystem changes"
SKOS	Simple Knowledge Organization System
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNIZAR	Universidad de Zaragoza
URI	Uniform Resource Identifier
WP5	EuroGEOSS Work Package 5 (Drought)

1 INTRODUCTION

This report aims at describing the Drought Vocabulary developed within EuroGEOSS WP5 and at documenting its creation process. This report is based on the article “The development and interlinkage of a drought vocabulary in the EuroGEOSS interoperable infrastructure” (Latre et al., 2011b), submitted on 2011-10-18, and currently under review, to the International Journal of Spatial Data Infrastructures Research.

This report is based on the work done in the thematic area of drought. The objectives of the drought working group are to connect drought-related resources on different spatial scales in an interoperable infrastructure (European Drought Observatory, EDO)[1]. The main elements of EDO are a drought metadata catalogue[2] for the discovery of drought-related data and services and a map viewer for visualizing drought indices.

Following the proposal of GEOSS and INSPIRE (European Commission, 2007) the discovery of information is based on searching through metadata descriptions. The drought team built a metadata catalogue that is tailored towards the needs of experts from the drought community. One of the key fields of the metadata describing resources is the field ‘keywords’ that facilitates the discovery of a resource of interest.

In the process of preparing metadata for drought-related data and services, it turned out that the proposed vocabularies within the EuroGEOSS project for the annotation of metadata did not comply with the needs of the drought community, because they were highly generic to allow for a proper classification of the resources or too large to be a practical tool for annotation.

It was decided to prepare a specialized drought vocabulary to improve the discovery of drought-related data and services in an interoperable infrastructure and to facilitate the task of metadata annotation. The resulting vocabulary, developed through an open and collaborative process, contains 103 concepts organized hierarchically in groups of concepts (drought, meteorology, soil, hydrology) and provides preferred and alternative labels in fifteen languages (Latre et al., 2011a).

The objective of this report is to describe the Drought Vocabulary and the methodology followed in its development (Lacasta et al., 2007).

The rest of the report is structured as follows. Section 2 reviews the state of the art in thesaurus related to drought and in thesaurus creation methodologies. Section 3 presents the drought vocabulary which was developed for a detailed annotation of metadata of drought related data and services. Section 4 focuses on the methodology followed for the development of the vocabulary. The integration of the vocabulary in the IOC is presented in section 5. The report finishes with a conclusions section (Section 6).

2 BACKGROUND AND RELATED WORK

A challenge in information retrieval from metadata catalogues is the provision of search results that are semantically related to the search terms. One approach to meet this challenge is the usage of controlled vocabularies, thesauri, or ontologies. An ontology is usually defined as “*a formal, explicit specification of a shared conceptualization*” (Gruber, 1993) and can be considered composed of a vocabulary of terms that refer to the things of interest in a given domain and some specification of

[1] http://edo.jrc.ec.europa.eu/
[2] http://eurogeoss.unizar.es/Search/

meaning for the terms (Uschold and Gruninger, 2004). When this specification is not fully specified by axioms and definitions; and just relationships among the terms (subtype/supertype, part/whole, synonym or relation) are made explicitly, these ontologies are usually referred to as thesauri or terminological ontologies (Lacasta et al., 2010; ISO, 1986; Sowa, 1996). In the case of multilingual thesauri, both terms and relationships are represented in more than one language. When applied to the search of resources, these multilingual thesauri allow the retrieval of resources that may not directly contain the search term among the annotation terms, but a term that is related to the search terms. This can be done by searching not only for the queried term, but also for the terms hierarchically dependant on it, and by the different translations the term may have (Latre et al., 2009). This has the advantage that the user retrieves a richer list of returns from his search.

2.1 Review of thematic thesauri related to drought

Two main issues were identified in the review of thesauri from fields close to drought and of thesauri proposed for metadata annotation in the EuroGEOSS project: they are either highly generic and contain only few terms related to drought or they are so extensive that their use for annotation of resources is impractical.

Thematic thesauri in fields close to drought, such as hydrography, hydrology and meteorology, are generally comprehensive; however, the amount of terms related to drought is limited. The Glossary of Meteorology[3] of the American Meteorological Society (2000) contains more than 12 000 terms related to meteorology and only a few are related to drought. The International Glossary of Hydrology (UNESCO, 1993), which counts with an experimental web version[4], is available in 11 languages and consists of more than 300 water-related terms, but few of them related to drought. The CUAHSI Water Ontology has the purpose of supporting the discovery of time-series data collected at a fixed point, including physical, chemical, and biological measurements. Again, with more than 5 000 terms, most of them not specific to drought, it is not practical for drought resources annotation and search. Extending the scope of the thesauri does not provide any improvement: AGROVOC[5], covering subject fields in agriculture, forestry and fisheries, contains close to 40 000 concepts, but only a dozen of them are drought-related.

General purpose thesauri, such as the thesauri proposed in the EuroGEOSS project for the annotation of metadata (INSPIRE topic categories, GEOSS Societal Benefit Areas categorization and the General Multilingual Environmental Thesaurus, GEMET) are not appropriate too for use in the drought field. Below an illustration of the issues related to the reuse of these thesauri:

- They can be highly generic vocabularies: INSPIRE topic categories (European Commission, 2008) and Societal Benefit Areas categorization[6] allow a categorization of data into general subject areas like 'climatologyMeteorologyAtmosphere' from the INSPIRE topic categories. These categories are too general to establish useful search restrictions by expert drought users when discovering data in a catalogue.

- They can be large collections of terms: GEMET[7] is a thesaurus containing around 65 000 terms. It is designed to cover a wide range of topic areas and the large amount of terms makes the selection of the right keywords for metadata annotation or restricting a discovery query tedious and cumbersome.

[5] http://aims.fao.org/website/AGROVOC-Thesaurus/sub
http://aims.fao.org/website/AGROVOC-Thesaurus/sub
[5] http://aims.fao.org/website/AGROVOC-Thesaurus/sub
[6] http://en.wikipedia.org/wiki/Societal_Benefit_Areas
[7] http://www.eionet.europa.eu/gemet

Since the need for a thesaurus for metadata annotation and improvement of searches in the drought metadata catalogue had been identified, this review of existing thesauri led to the preparation of a specialized vocabulary on droughts.

2.2 State of art in the process of thesaurus construction

The construction of a thesaurus is a complex process in which the terminology used in a knowledge domain is collected, analyzed and linked together into a model that can be used for classification of resources in the domain. Along the years, with the objective of improving the quality of the created models, different thesaurus construction methodologies have been developed. In this field, different standards have been created to normalize the structure and properties of monolingual and multilingual thesauri (ANSI/NISO, 2005; ISO, 2011; BSI, 2007). These standards do not propose a detailed construction methodology, but they describe the general idea of the most common processes used for thesauri construction. In general four steps are usually required:

- A review of similar existent thesauri. This is needed to avoid the creation of a new thesaurus from zero if an existent one can be valid or adapted.

- A modelling stage where the desired structure, format and final display are selected.

- A term selection stage where the set of possible terms to include in the thesaurus are selected and related.

- A validation step in which the candidate terms are reviewed to select only those that fulfil the standards specifications.

Depending on the specific methodology used, each one of these steps can be performed in a different way. For example, the term selection stage can be performed by a committee generating a corpus of terms or they can be extracted from the domain (e.g., other existent knowledge models). And in each of these cases different approaches can be used. In the first case, the corpus can be constructed from general terms to specific ones (top-down) or vice versa (bottom-up). In the second one, each extracted term can be directly used in the model (inductive approach) or reviewed after the whole desired terms is extracted (deductive approach).

Following these general guidelines, De Vorsey et al. (2006) describe the process used to construct the American Museum of Natural History (AMNH) Thesaurus. The process starts with the revision of existent thesauri in the area and then uses a subset of Art & Architecture Thesaurus (AAT) and a set of keywords already used in the AMNH as candidate terms for the thesaurus. The process includes a cleaning phase of the AAT, a second one of harmonization of AMNH keywords and a third one of definition of relations and scope notes for the selected terms.

Other works provide construction methodologies partially different from the indicated by thesaurus standards. For example, the State Records Authority of New South Wales (2003) describes a complete thesaurus construction process whose term collection phase is based on DIRKS methodology (Commonwealth of Australia, 2001) for the construction of the organization classification schemes. The complete methodology has a first stage of preparation (review of thesaurus need and planning); a second one devoted to collecting information that uses the DIRKS methodology and interviews to future users; a third one of analysis, where the thesaurus structure is composed; a fourth one of collation where the model is represented in a final format; a fifth one of revision, where feedback is searched; and a final one of production where the created thesaurus is put into use.

The process described by the Semantic Health Project (2006) to create the Belgian Bilingual Bi-encoded Thesaurus (3BT) is less elaborated. It uses the Amsterdam Thesaurus (AT) as an initial version and then it removes the unrequired concepts. Finally, it applies a set of refinement stages that add concepts, correct linguistic errors and translate the keywords from the German and French teams. The Commonwealth of Australia (2003) also describes a thesaurus construction process in a quite general way. It assumes that the organization has developed a business classification scheme in accordance with the DIRKS methodology. And then it provides an eight-step guide to convert such scheme into a thesaurus. Another approach is the one indicated by the Working Group on Guidelines for Multilingual Thesauri of the IFLA Classification and Indexing Section (2005), which describes a methodology for the construction of multilingual thesaurus (from scratch and in base to others). However, it focuses on the criteria for selection of symmetrical terms in different languages, not on the basics for selecting terms and identifying relationships.

Finally, there are approaches based on techniques used for the construction of ontologies. In this context, Bechhofer and Goble (2001) describe a construction process that use knowledge representation techniques to facilitate the construction of coherent hierarchies. It does not describe a proper methodology, but it uses the bottom-up approach described by Vickery (1966) and improves it using Description Logics as the scheme to model the relationships between the concepts more precisely.

3 DROUGHT VOCABULARY DESCRIPTION

The drought team of the EuroGEOSS project prepared a metadata catalogue tailored towards users from the drought community with data and services linked to the drought field. One of the key fields of the metadata describing resources is the field ‘keywords’, that facilitates the discovery of a resource of interest. Since neither the proposed general purpose vocabularies within the EuroGEOSS for the annotation of metadata nor the hydrology or meteorology related ones did comply with the needs of the drought community, it was decided to prepare a specialized drought vocabulary to:

- tailor the search to drought specific content,
- support the data providers in the annotation task of the metadata,
- approach the issue of dealing with search terms in various languages.

The resulting vocabulary, developed through an open and collaborative process, contains 103 concepts organized hierarchically in groups of concepts (drought, meteorology, soil, hydrology) and provides preferred and alternative labels in fifteen languages (Latre et al., 2011a,b).

Terms and relations of the vocabulary are shown with English labels below in Figure 1 to Figure 6. A full PDF file showing together all the concepts and relations can be downloaded from http://eurogeoss.unizar.es/home/thesaurus/droughtVocabulary.pdf.

Figure 1: Graphical representation of part of the drought vocabulary: drought group

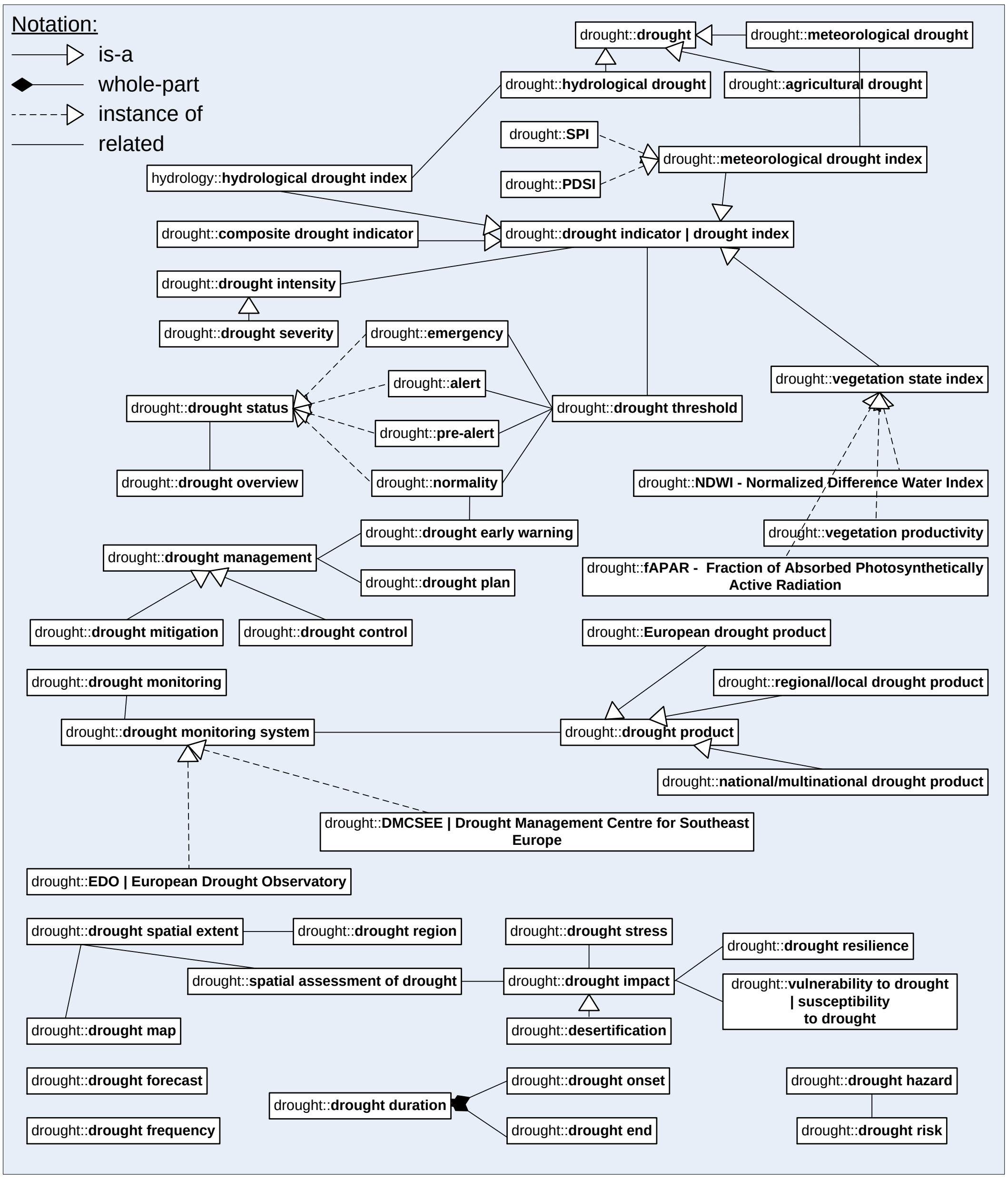

Notation:
is-a
whole-part
instance of
related
drought::drought
drought::meteorological drought
drought::hydrological drought
drought::agricultural drought
drought::SPI
drought::meteorological drought index
drought::PDSI
hydrology::hydrological drought index
drought::composite drought indicator
drought::drought indicator | drought index
drought::drought intensity
drought::drought severity
drought::emergency
drought::alert
drought::drought status
drought::drought threshold
drought::pre-alert
drought::vegetation state index
drought::drought overview
drought::normality
drought::NDWI - Normalized Difference Water Index
drought::drought early warning
drought::vegetation productivity
drought::drought management
drought::fAPAR - Fraction of Absorbed Photosynthetically Active Radiation
drought::drought plan
drought::drought mitigation
drought::drought control
drought::European drought product
drought::drought monitoring
drought::regional/local drought product
drought::drought monitoring system
drought::drought product
drought::national/multinational drought product
drought::DMCSEE | Drought Management Centre for Southeast Europe
drought::EDO | European Drought Observatory
drought::drought spatial extent
drought::drought region
drought::drought stress
drought::drought resilience
drought::spatial assessment of drought
drought::drought impact
drought::vulnerability to drought | susceptibility to drought
drought::drought map
drought::desertification
drought::drought forecast
drought::drought onset
drought::drought hazard
drought::drought duration
drought::drought frequency
drought::drought end
drought::drought risk

Figure 2: Graphical representation of part of the drought vocabulary: meteorology group

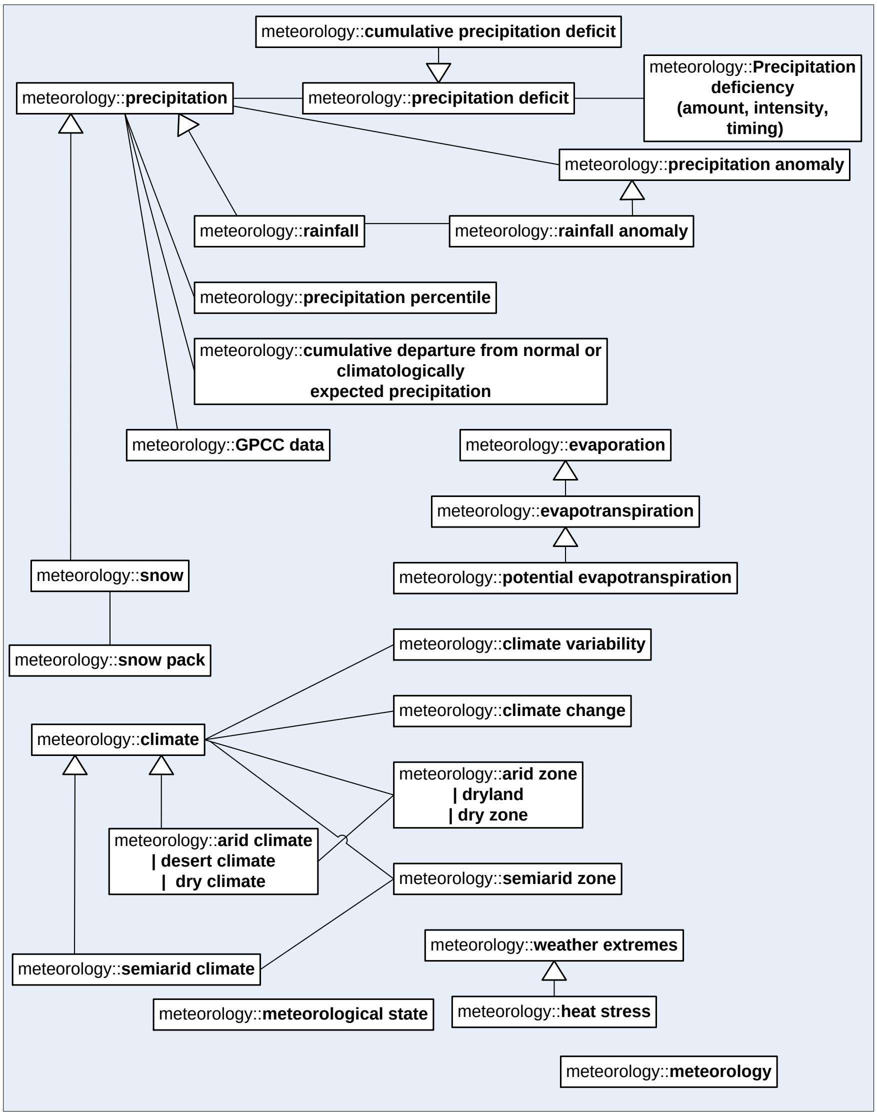

meteorology::cumulative precipitation deficit
meteorology::precipitation
meteorology::precipitation deficit
meteorology::Precipitation deficiency (amount, intensity, timing)
meteorology::precipitation anomaly
meteorology::rainfall
meteorology::rainfall anomaly
meteorology::precipitation percentile
meteorology::cumulative departure from normal or climatologically expected precipitation
meteorology::GPCC data
meteorology::evaporation
meteorology::evapotranspiration
meteorology::snow
meteorology::potential evapotranspiration
meteorology::snow pack
meteorology::climate variability
meteorology::climate change
meteorology::climate
meteorology::arid zone | dryland | dry zone
meteorology::arid climate | desert climate | dry climate
meteorology::semiarid zone
meteorology::weather extremes
meteorology::semiarid climate
meteorology::meteorological state
meteorology::heat stress
meteorology::meteorology

Figure 3: Graphical representation of part of the drought vocabulary: hydrology group

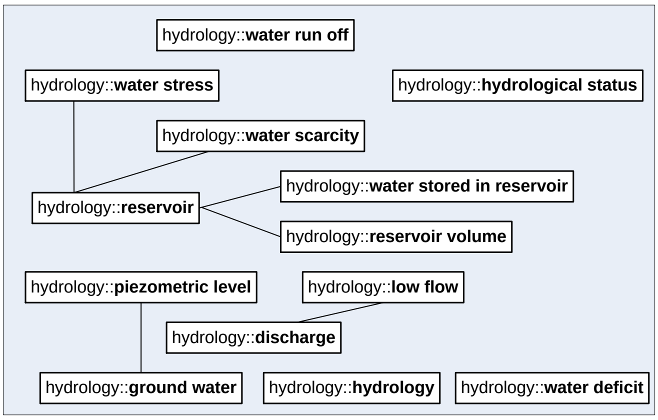


Figure 4: Graphical representation of part of the drought vocabulary: soil group

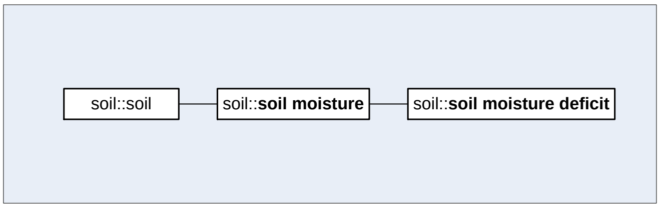


Figure 5: Graphical representation of part of the drought vocabulary: statistics group

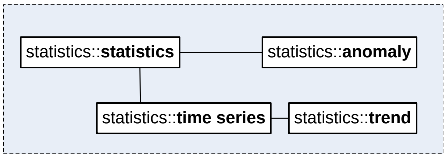
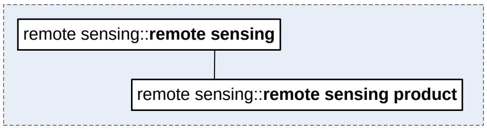


Figure 6: Graphical representation of part of the drought vocabulary: remote sensing group

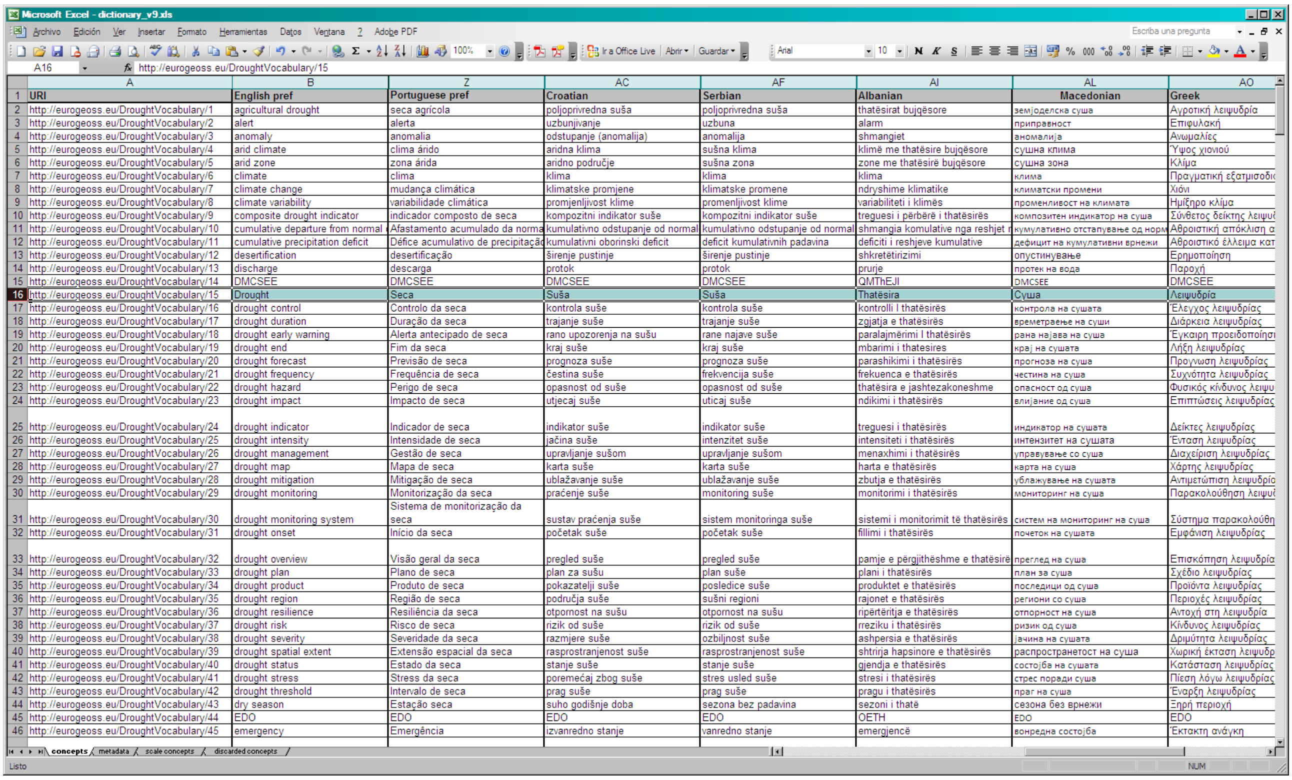

remote sensing::**remote sensing**
remote sensing::**remote sensing product**

The Drought Vocabulary has been translated into 15 languages:

- English
- Slovenian
- Spanish
- French
- German
- Bosnian
- Turkish
- Italian
- Portuguese
- Croatian
- Serbian
- Albanian
- Macedonian
- Greek
- Montenegrin

Figure 7 shows a preview of the spreadsheet with the labels of the terms belonging to the Drought Vocabulary in the fifteen different languages, that can be downloaded from http://eurogeoss.unizar.es/home/thesaurus/EuroGEOSS_Drought_Vocabulary_labels.xls.

Figure 7: Spreadsheet with the labels of the terms belonging to the Drought Vocabulary in the fifteen different languages


Microsoft Excel - dictionary_v9.xls
Archivo Edición Ver Insertar Formato Herramientas Datos Ventana ? Adobe PDF
Escriba una pregunta
100%
Ir a Office Live
Abrir
Guardar
Arial
10
A16
http://eurogeoss.eu/DroughtVocabulary/15
concepts
metadata
scale concepts
discarded concepts
Listo
NUM

	A	B	Z	AC	AF	AI	AL	AO
1	URI	English pref	Portuguese pref	Croatian	Serbian	Albanian	Macedonian	Greek
2	http://eurogeoss.eu/DroughtVocabulary/1	agricultural drought	seca agrícola	poljoprivredna suša	poljoprivredna suša	thatësirat bujqësore	земјоделска суша	Αγροτική λειψυδρία
3	http://eurogeoss.eu/DroughtVocabulary/2	alert	alerta	uzbunjivanje	uzbuna	alarm	приправност	Επιφυλακή
4	http://eurogeoss.eu/DroughtVocabulary/3	anomaly	anomalia	odstupanje (anomalija)	anomalija	shmangiet	аномалија	Ανωμαλίες
5	http://eurogeoss.eu/DroughtVocabulary/4	arid climate	clima árido	aridna klima	sušna klima	klimë me thatësire bujqësore	сушна клима	Ύψος χιονιού
6	http://eurogeoss.eu/DroughtVocabulary/5	arid zone	zona árida	aridno područje	sušna zona	zone me thatësirë bujqësore	сушна зона	Κλίμα
7	http://eurogeoss.eu/DroughtVocabulary/6	climate	clima	klima	klima	klima	клима	Πραγματική εξατμισοδι
8	http://eurogeoss.eu/DroughtVocabulary/7	climate change	mudança climática	klimatske promjene	klimatske promene	ndryshime klimatike	климатски промени	Χιόνι
9	http://eurogeoss.eu/DroughtVocabulary/8	climate variability	variabilidade climática	promjenljivost klime	promenljivost klime	variabiliteti i klimës	променливост на климата	Ημίξηρο κλίμα
10	http://eurogeoss.eu/DroughtVocabulary/9	composite drought indicator	indicador composto de seca	kompozitni indikator suše	kompozitni indikator suše	treguesi i përbërë i thatësirës	композитен индикатор на суша	Σύνθετος δείκτης λειψυδ
11	http://eurogeoss.eu/DroughtVocabulary/10	cumulative departure from normal	Afastamento acumulado da norma	kumulativno odstupanje od normal	kumulativno odstupanje od normal	shmangia komulative nga reshjet r	кумулативно отстапување од норм	Αθροιστική απόκλιση α
12	http://eurogeoss.eu/DroughtVocabulary/11	cumulative precipitation deficit	Défice acumulativo de precipitaçã	kumulativni oborinski deficit	deficit kumulativnih padavina	deficiti i reshjeve kumulative	дефицит на кумулативни врнежи	Αθροιστικό έλλειμα κατ
13	http://eurogeoss.eu/DroughtVocabulary/12	desertification	desertificação	širenje pustinje	širenje pustinje	shkretëtirizimi	опустинување	Ερημοποίηση
14	http://eurogeoss.eu/DroughtVocabulary/13	discharge	descarga	protok	protok	prurje	протек на вода	Παροχή
15	http://eurogeoss.eu/DroughtVocabulary/14	DMCSEE	DMCSEE	DMCSEE	DMCSEE	QMThEJI	DMCSEE	DMCSEE
16	http://eurogeoss.eu/DroughtVocabulary/15	Drought	Seca	Suša	Suša	Thatësira	Суша	Λειψυδρία
17	http://eurogeoss.eu/DroughtVocabulary/16	drought control	Controlo da seca	kontrola suše	kontrola suše	kontrolli I thatësirës	контрола на сушата	Έλεγχος λειψυδρίας
18	http://eurogeoss.eu/DroughtVocabulary/17	drought duration	Duração da seca	trajanje suše	trajanje suše	zgjatja e thatësirës	времетраење на суши	Διάρκεια λειψυδρίας
19	http://eurogeoss.eu/DroughtVocabulary/18	drought early warning	Alerta antecipado de seca	rano upozorenja na sušu	rane najave suše	paralajmërimi I thatësirës	рана најава на суша	Έγκαιρη προειδοποίησ
20	http://eurogeoss.eu/DroughtVocabulary/19	drought end	Fim da seca	kraj suše	kraj suše	mbarimi i thatesires	крај на сушата	Λήξη λειψυδρίας
21	http://eurogeoss.eu/DroughtVocabulary/20	drought forecast	Previsão de seca	prognoza suše	prognoza suše	parashikimi i thatësirës	прогноза на суша	Προγνωση λειψυδρίας
22	http://eurogeoss.eu/DroughtVocabulary/21	drought frequency	Frequência de seca	čestina suše	frekvencija suše	frekuenca e thatësirës	честина на суша	Συχνότητα λειψυδρίας
23	http://eurogeoss.eu/DroughtVocabulary/22	drought hazard	Perigo de seca	opasnost od suše	opasnost od suše	thatësira e jashtezakoneshme	опасност од суша	Φυσικός κίνδυνος λειψυ
24	http://eurogeoss.eu/DroughtVocabulary/23	drought impact	Impacto de seca	utjecaj suše	uticaj suše	ndikimi i thatësirës	влијание од суша	Επιπτώσεις λειψυδρίας
25	http://eurogeoss.eu/DroughtVocabulary/24	drought indicator	Indicador de seca	indikator suše	indikator suše	treguesi i thatësirës	индикатор на сушата	Δείκτες λειψυδρίας
26	http://eurogeoss.eu/DroughtVocabulary/25	drought intensity	Intensidade de seca	jačina suše	intenzitet suše	intensiteti i thatësirës	интензитет на сушата	Ένταση λειψυδρίας
27	http://eurogeoss.eu/DroughtVocabulary/26	drought management	Gestão de seca	upravljanje sušom	upravljanje sušom	menaxhimi i thatësirës	управување со суша	Διαχείριση λειψυδρίας
28	http://eurogeoss.eu/DroughtVocabulary/27	drought map	Mapa de seca	karta suše	karta suše	harta e thatësirës	карта на суша	Χάρτης λειψυδρίας
29	http://eurogeoss.eu/DroughtVocabulary/28	drought mitigation	Mitigação de seca	ublažavanje suše	ublažavanje suše	zbutja e thatësirës	ублажување на сушата	Αντιμετώπιση λειψυδρία
30	http://eurogeoss.eu/DroughtVocabulary/29	drought monitoring	Monitorização da seca	praćenje suše	monitoring suše	monitorimi i thatësirës	мониторинг на суша	Παρακολούθηση λειψυδ
31	http://eurogeoss.eu/DroughtVocabulary/30	drought monitoring system	Sistema de monitorização da seca	sustav praćenja suše	sistem monitoringa suše	sistemi i monitorimit të thatësirës	систем на мониторинг на суша	Σύστημα παρακολούθη
32	http://eurogeoss.eu/DroughtVocabulary/31	drought onset	Início da seca	početak suše	početak suše	fillimi i thatësirës	почеток на сушата	Εμφάνιση λειψυδρίας
33	http://eurogeoss.eu/DroughtVocabulary/32	drought overview	Visão geral da seca	pregled suše	pregled suše	pamje e përgjithëshme e thatësirë	преглед на суша	Επισκόπηση λειψυδρία
34	http://eurogeoss.eu/DroughtVocabulary/33	drought plan	Plano de seca	plan za sušu	plan suše	plani i thatësirës	план за суша	Σχέδιο λειψυδρίας
35	http://eurogeoss.eu/DroughtVocabulary/34	drought product	Produto de seca	pokazatelji suše	posledice suše	produktet e thatësirës	последици од суша	Προϊόντα λειψυδρίας
36	http://eurogeoss.eu/DroughtVocabulary/35	drought region	Região de seca	područja suše	sušni regioni	rajonet e thatësirës	региони со суша	Περιοχές λειψυδρίας
37	http://eurogeoss.eu/DroughtVocabulary/36	drought resilience	Resiliência da seca	otpornost na sušu	otpornost na sušu	ripërtëritja e thatësirës	отпорност на суша	Αντοχή στη λειψυδρία
38	http://eurogeoss.eu/DroughtVocabulary/37	drought risk	Risco de seca	rizik od suše	rizik od suše	rreziku i thatësirës	ризик од суша	Κίνδυνος λειψυδρίας
39	http://eurogeoss.eu/DroughtVocabulary/38	drought severity	Severidade da seca	razmjere suše	ozbiljnost suše	ashpersia e thatësirës	јачина на сушата	Δριμύτητα λειψυδρίας
40	http://eurogeoss.eu/DroughtVocabulary/39	drought spatial extent	Extensão espacial da seca	rasprostranjenost suše	rasprostranjenost suše	shtrirja hapsinore e thatësirës	распространетост на суша	Χωρική έκταση λειψυδρ
41	http://eurogeoss.eu/DroughtVocabulary/40	drought status	Estado da seca	stanje suše	stanje suše	gjendja e thatësirës	состојба на сушата	Κατάσταση λειψυδρίας
42	http://eurogeoss.eu/DroughtVocabulary/41	drought stress	Stress da seca	poremećaj zbog suše	stres usled suše	stresi i thatësirës	стрес поради суша	Πίεση λόγω λειψυδρίας
43	http://eurogeoss.eu/DroughtVocabulary/42	drought threshold	Intervalo de seca	prag suše	prag suše	pragu i thatësirës	праг на суша	Έναρξη λειψυδρίας
44	http://eurogeoss.eu/DroughtVocabulary/43	dry season	Estação seca	suho godišnje doba	sezona bez padavina	sezoni i thatë	сезона без врнежи	Ξηρή περιοχή
45	http://eurogeoss.eu/DroughtVocabulary/44	EDO	EDO	EDO	EDO	OETH	EDO	EDO
46	http://eurogeoss.eu/DroughtVocabulary/45	emergency	Emergência	izvanredno stanje	vanredno stanje	emergjencë	вонредна состојба	Έκτακτη ανάγκη

The Drought Vocabulary has been formalized in SKOS (Miles and Pérez-Agüera, 2007). SKOS (Simple Knowledge Organization System) is a family of formal languages designed for representation of thesauri, classification schemes, taxonomies, subject-heading systems, or any other type of structured controlled vocabulary. Built upon RDF (Resource Description Format) and RDFS (RDF Schema), its main objective is to enable easy publication of controlled structured vocabularies for the Semantic Web. Unique URIs (Uniform Resource Identifier) were constructed for the terms (such as '*http://eurogeoss.eu/DroughtVocabulary/15*' for 'drought') in order to allow referring to a term in a language-independent manner. The *is-a*, *whole-part* and *instance-of* relationships were mapped to the *skos:broader* and *skos:narrower* relationships while the *related* relationship has been maintained too. There exists also the possibility of grouping concepts using the *skos:collections* construction to provide a more consistent grouping of the terms into the different categories identified during the modelling stage: meteorology, drought, soil, hydrology, statistics. The SKOS version of the vocabulary is available at http://eurogeoss.unizar.es/home/thesaurus/DroughtVocabulary.skos.xml.

Finally, its Dublin Core metadata can be downloaded as an XML file from http://eurogeoss.unizar.es/home/thesaurus/DroughtVocabulary.MD.DC.xml.

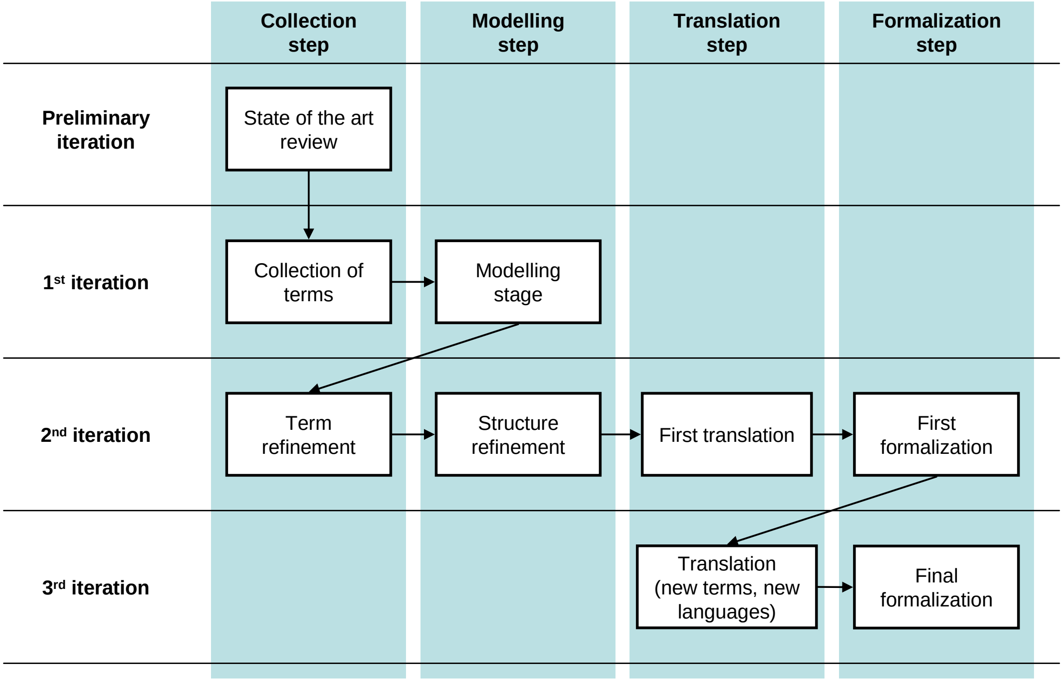
4 DEVELOPMENT OF THE VOCABULARY

4.1 Overview of the methodology

The methodology followed for the development of the drought vocabulary is quite similar to the one described in the ISO standards (see section 2.2). It includes a review, modelling and structure refinement stages. However, these steps were applied in an iterative way and an additional formalization step was included to be able to use the thesaurus in an information retrieval environment. In the context of the drought work package of EuroGEOSS, this iterative process allowed for a rapid integration in the technological infrastructure being developed as part of the first tasks of the project and it also facilitated quick feedback and flexible collaboration of the different partners in the development and refinement of the vocabulary. A total of three iterations were needed to create the vocabulary. Each iteration can include steps of information collection, modelling, translation and formalization with a different degree of emphasis. Previously to the first iteration, a state of the art review was performed to ensure that no other thesaurus or vocabulary fitted our purposes. Most thesauri proved to be too general or to big to be used in the drought field, as explained in section 2.

The main part of the effort devoted to the creation of the vocabulary was made during the first iteration. The second one was dedicated to refine the vocabulary based on the results of the first iteration and to provide a first translation of the terms and a draft formalization in a knowledge representation language. The third and final iteration was devoted to finish the translations and to obtain the final formalized version of the vocabulary. Figure 8 shows a schema of the different steps and iterations followed to create the thesaurus, that are explained in the following sections.

Figure 8. Steps of the methodology followed in the drought vocabulary creation



4.2 Collection of information

The initial step of the first and second iterations was the selection of terms. During the first iteration, information and terms were collected: all partners of the drought work package contributed a list of keywords in a common language (English) that described their data and services. In most cases, the submitted terms have already been used informally to tag the created metadata. Apart from the knowledge from partner experts, related terms in well-known sources (GEMET and AGROVOC thesauri) were also searched. The final set of keywords contained terms that allow characterizing drought events, drought data, general topics related to droughts, drought indices, etc. The initial list was refined in the second iteration in order to add missing terms, (such as ‘drought risk’, ‘drought management plan’, ‘discharge’, ‘drought resilience’) or prune not very related ones (such as terms describing time and spatial scale, since there are more specific thesauri to cover that).

4.3 Modelling

After the collection of terms, a modelling stage took place. An identification of synonyms or conflation of different submitted keywords referring to the same term was made in the first iteration. Preferred and alternate labels were then chosen among the keywords for the conflated terms. For instance, 'arid climate', 'desert climate' and 'dry climate' were different keywords individually proposed by different partners, all of them referring to the same term. 'Arid climate' was chosen as preferred label and the other two were maintained as alternate labels. In the case of the terms referred by acronyms or the pairs 'acronym-complete name', terms were split in two in order to separate acronyms (preferred label) from their complete name (alternate label). Finally, a draft structure or hierarchy for the terms was proposed.

In the second iteration, this modelling was refined: different categories (groups of concepts) were identified and the hierarchical relationships of the first version were refined, maintaining only as purely hierarchical those that could be classified into *is-a* relationships (a 'rainfall anomaly' is a kind of 'precipitation anomaly'), *whole-part* ('drought duration' has an 'onset' and an 'end') or *instance-of* ('EDO' is an instance/individual/particular case of a 'drought monitoring system'). The rest were considered as non-hierarchical and maintained as simple *related* relationships ('soil' is related to 'soil moisture').

4.4 Translation

The third step in the creation of the thesaurus was the translation of the terms. For the first version of the vocabulary, obtained at the end of the second iteration, translations into Slovenian, Spanish, French and German, besides the original English version, could be quickly provided by the partners and, thus, available for testing multilingualism aspects of the use of the thesaurus. In the third iteration, apart from the translation of the newly added terms, translations into Bosnian, Turkish, Italian, Portuguese, Croatian, Serbian, Albanian, Macedonian, Greek and Montenegrin were integrated into the thesaurus, to sum up a total of 15 languages.

4.5 Formalization

The final step in the creation of the thesaurus was its formalization. The first version of the thesaurus was represented in SKOS for testing purposes. The final version has been integrated into the JRC SESAME repository and is accessible from the EuroGEOSS Drought Catalogue home page[8]

5 INTEGRATION OF THE THESAURUS IN THE EUROGEOSS FRAMEWORK FOR DROUGHT MONITORING

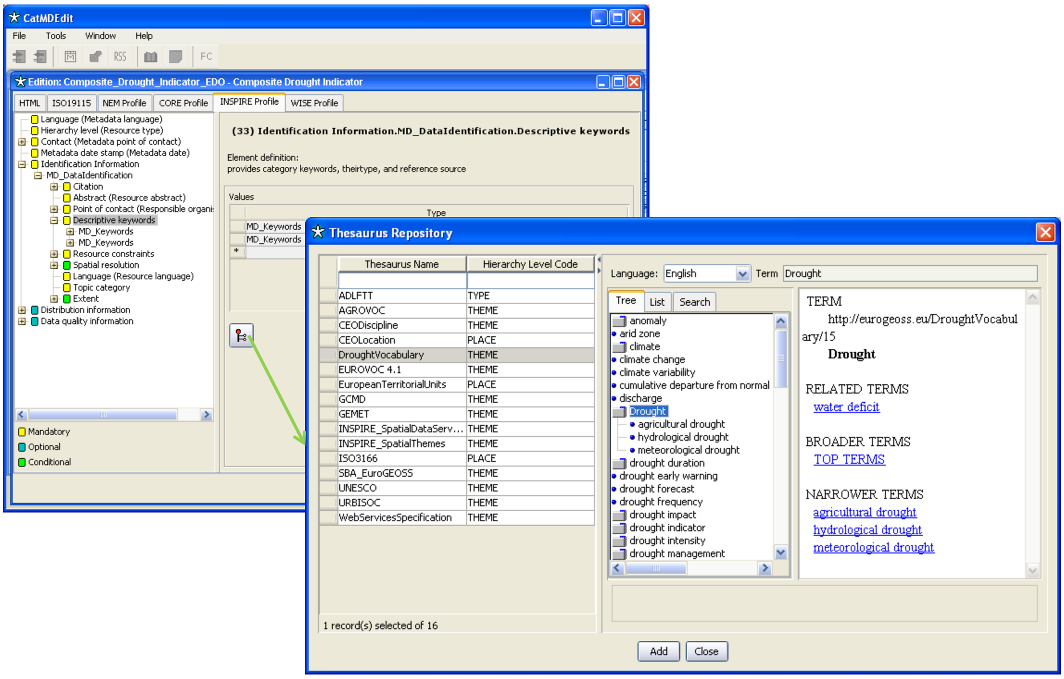
The drought vocabulary has been integrated in the infrastructure of the European Drought Observatory in three ways. Firstly, it has been incorporated into the CatMDEdit tool, the metadata editor tool used in the drought working group of the EuroGEOSS project. Secondly, it has been integrated within the web application used for searching and updating online the metadata records. And thirdly, has been aligned with the other two thesauri used in the EuroGEOSS framework: GEMET and the GEOSS Societal Benefit Areas categories.

[8] http://eurogeoss.unizar.es/home/

5.1 Integration into the EuroGEOSS drought metadata editor tool for resource annotation

The metadata editor tool used in the EuroGEOSS drought work package, CatMDEdit[9] (Nogueras-Iso et al., 2008), uses a serialized version of the vocabulary to browse its content and allows users creating or updating metadata to select terms from the vocabulary to tag the resources. Figure 9 shows the drought vocabulary in the CatMDEdit thesaurus browser. The vocabulary can be browsed through the thesaurus treelike structure or through an alphabetical term list. Additionally, there is a third tab that allows searching terms contained in the thesaurus. The thesaurus browser, loaded with the drought vocabulary, allows users to semantically annotate the resources with selected concepts from the vocabulary.

Figure 9: Integration of the EuroGEOSS Drought Vocabulary in CatMDEdit



- Language (Metadata language)
- Hierarchy level (Resource type)
- Contact (Metadata point of contact)
- Metadata date stamp (Metadata date)
- Identification Information
  - MD_DataIdentification
    - Citation
    - Abstract (Resource abstract)
    - Point of contact (Responsible organi:
    - Descriptive keywords
      - MD_Keywords
      - MD_Keywords
    - Resource constraints
    - Spatial resolution
    - Language (Resource language)
    - Topic category
    - Extent
- Distribution information
- Data quality information

🟨 Mandatory
🟦 Optional
🟩 Conditional

(33) Identification Information.MD_DataIdentification.Descriptive keywords

Element definition:
provides category keywords, theirtype, and reference source

Values

Type
MD_Keywords
MD_Keywords
*

Thesaurus Name	Hierarchy Level Code
ADLFTT	TYPE
AGROVOC	THEME
CEODiscipline	THEME
CEOLocation	PLACE
DroughtVocabulary	THEME
EUROVOC 4.1	THEME
EuropeanTerritorialUnits	PLACE
GCMD	THEME
GEMET	THEME
INSPIRE_SpatialDataServ…	THEME
INSPIRE_SpatialThemes	THEME
ISO3166	PLACE
SBA_EuroGEOSS	THEME
UNESCO	THEME
URBISOC	THEME
WebServicesSpecification	THEME

1 record(s) selected of 16

Language: English Term Drought

- anomaly
- arid zone
- climate
- climate change
- climate variability
- cumulative departure from normal
- discharge
- Drought
  - agricultural drought
  - hydrological drought
  - meteorological drought
- drought duration
- drought early warning
- drought forecast
- drought frequency
- drought impact
- drought indicator
- drought intensity
- drought management

TERM
http://eurogeoss.eu/DroughtVocabulary/15
Drought

RELATED TERMS
water deficit

BROADER TERMS
TOP TERMS

NARROWER TERMS
agricultural drought
hydrological drought
meteorological drought

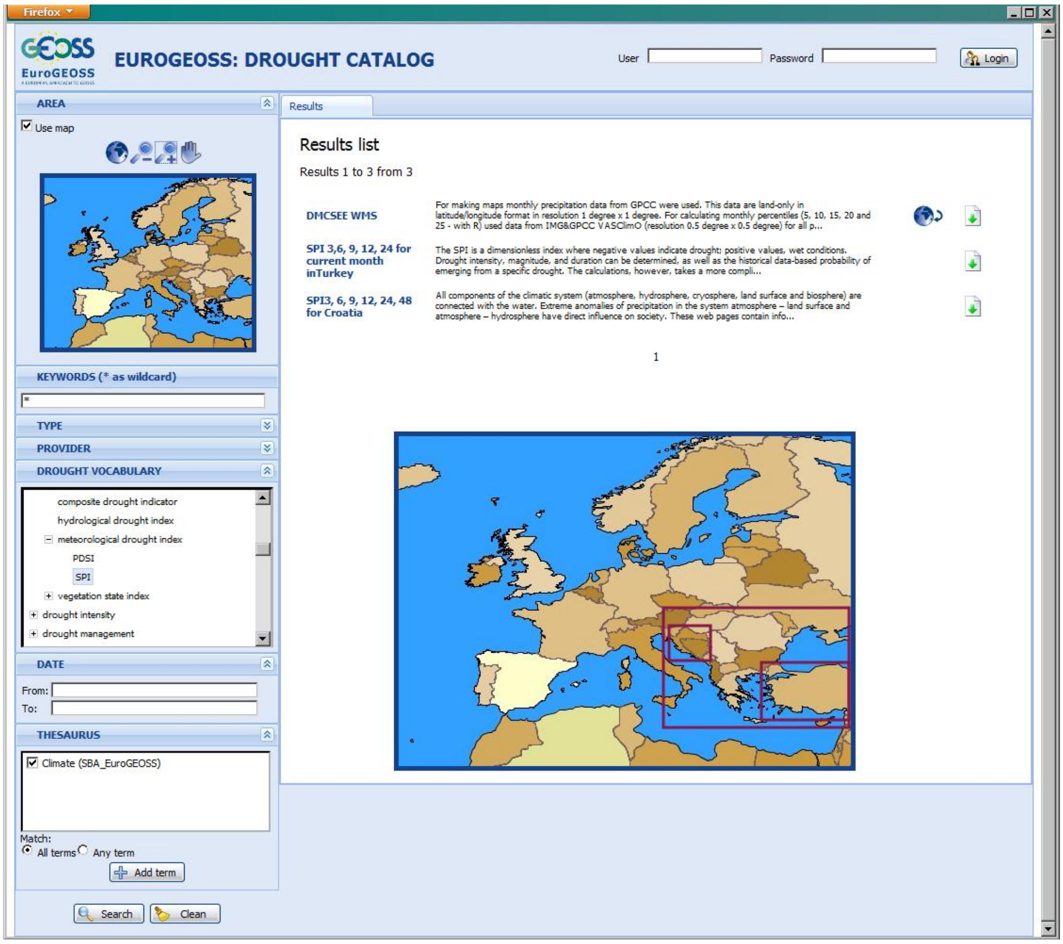
5.2 Integration into EuroGEOSS drought catalogue user interface

The drought vocabulary has also been integrated into the drought catalogue user interface. The metadata records managed by this catalogue are ISO 19115 and INSPIRE compliant and describe 210 datasets and 22 web services submitted by the EuroGEOSS drought partners. About a 58% of the records were written in the original language of the dataset, while the rest were in English, what made difficult the discovery of the described resources when querying by a different language.

The web catalogue is accessible through an OGC compliant catalogue service developed with CatalogCube technology[10] and through a user friendly web application. This web application was designed to take into account the GCI Clearinghouse Requirements (GEO, 2009) about searching criteria: users should be able to search based on location, keywords or text, and temporal extent. In addition to this, a *resource type* and a *provider* criterion were also included as a way to allow users to distinguish the resource type (data or services) or the resource provider in their searches. Figure 10 shows this GUI covering these searching criteria.

[9] http://catmdedit.sourceforge.net/
[10] http://spatiumcube.sourceforge.net/

Figure 10: Graphical user interface web application for searching metadata



☑ Use map

Keywords: *

composite drought indicator
hydrological drought index
⊟ meteorological drought index
PDSI
SPI
⊞ vegetation state index
⊞ drought intensity
⊞ drought management

From:
To:

☑ Climate (SBA_EuroGEOSS)

Match:
● All terms ○ Any term

Results list

Results 1 to 3 from 3

DMCSEE WMS	For making maps monthly precipitation data from GPCC were used. This data are land-only in latitude/longitude format in resolution 1 degree x 1 degree. For calculating monthly percentiles (5, 10, 15, 20 and 25 - with R) used data from IMG&GPCC VASClimO (resolution 0.5 degree x 0.5 degree) for all p...
SPI 3,6, 9, 12, 24 for current month inTurkey	The SPI is a dimensionless index where negative values indicate drought; positive values, wet conditions. Drought intensity, magnitude, and duration can be determined, as well as the historical data-based probability of emerging from a specific drought. The calculations, however, takes a more compli...
SPI3, 6, 9, 12, 24, 48 for Croatia	All components of the climatic system (atmosphere, hydrosphere, cryosphere, land surface and biosphere) are connected with the water. Extreme anomalies of precipitation in the system atmosphere – land surface and atmosphere – hydrosphere have direct influence on society. These web pages contain info...

1

Once the first version of the drought vocabulary was developed, it was displayed in the interface in order to facilitate querying the catalogue by using drought-specific related terms. It took the place of the thesauri that were in use at the moment (INSPIRE and SBA categories and sub-categories), since they had proved to be too general to aid the searching. The drought vocabulary, even in its first version and prior to the updating of the metadata with tags from the new vocabulary, proved to be an improvement in the interaction with the user, due to its capability of describing better the resources and the fact that its terms were already used informally in the keywords, abstract or title sections of the resources. Section **Erreur ! Source du renvoi introuvable.** discusses this in more depth.

5.3 Aligment with other thesauri

The Drought Vocabulary is tailored towards optimizing searches in the drought metadata catalogue. To make the vocabulary useful also in a wider context, it needs to be linked with thesauri that are used for searching through multidisciplinary metadata catalogues. In the EuroGEOSS project, these thesauri are GEMET and GEOSS Societal Benefit Areas (SBA). The linking of thesauri is referred to as matching: all terms of the drought vocabulary have to be matched to at least one term of another thesaurus. The process needs to be repeated for every thesaurus that needs to the linked to the drought vocabulary. The matching was performed manually with the SKOS matcher of the Semantic Lab[11] of the Joint Research Centre.

[11] http://semanticlab.jrc.ec.europa.eu/

A summary of the alignment activity between the drought vocabulary and either the SBA or GEMET is presented in the following tables. Table 1 shows that 12 concepts out of the 66 of the SBA were mapped to 45 concepts of the drought vocabulary, and that 50 terms out of 5244 from GEMET were mapped to a total of 103 concepts of the drought vocabulary. This validates the authors' claim made in section 2.1 that the general purpose thesauri were too big or too general for its use in metadata annotation and search in the drought area.
Table 2 also justifies the need of a drought vocabulary: most of the mappings between terms of the drought vocabulary and the considered thesauri belong to the category of "related terms". Only 5 out of 125 mappings in the case of the SBA and 23 out of 137 in the case of GEMET are more specific mappings (broader, close and exact matches).

Table 1. Number of mapped concepts

	Thesaurus size	# of mapped concepts	
		from original thesaurus	to the drought vocabulary
SBA	66	12	45
GEMET	5244	50	103

Table 2. Number of mapping relations to the drought vocabulary

	Broad	Close	Exact	Related	Total
SBA	2	1	2	120	125
GEMET	8	1	14	114	137

The big advantage of the matching of vocabularies is that the search of the user can be automatically extended to terms of the drought vocabulary that are linked to the selected term of the GEMET or GEOSS SBA.

6 CONCLUSIONS

The proposed vocabularies in the European project EuroGEOSS for the annotation of metadata proven to be, in the thematic area of drought, either too generic to adequately classify drought resources, or too large to be practical for their annotation. As a consequence, a drought vocabulary has been developed in a collective way, in order to improve the accessibility to appropriate drought resources (datasets and services) to users and experts.

It was thought that a first-guess vocabulary could be prepared for this area based on a collection of terms that would considerably improved the discovery of available resources and, in the end, a 103-term vocabulary, organized into a hierarchy and translated into 15 languages has been developed. The methodology followed for the creation of this specific drought vocabulary has been presented, methodology that could be also applied to other subject areas where the same needs and problems could be identified.

This vocabulary has been first used in the EuroGEOSS drought catalogue in three ways. Firstly, by annotating the resources it holds according to the new vocabulary, since the quality of the search results of a catalogue query depends on the quality of the metadata. The terms of the vocabulary had to be used in the annotation of the metadata, since otherwise resources cannot be properly found. A 51.5% of the drought vocabulary concepts have been used in the annotation of the EuroGEOSS drought catalogue resources. Secondly, it has been included in the user search application interface. Analysis from the catalogue logs shows that it improves the interaction with users, helping them to establish their searching parameters and providing them with better search results (Latre et al. 2011b). Logs also show that the terms selected to be part of the vocabulary are appropriate from a user's point of view, since 65.0% of them have been used in at least a query.

Finally, the vocabulary has been aligned with the other two thesauri chosen for metadata annotation in the project: GEMET and GEOSS Societal Benefit Areas categories.

As future work, the Drought Vocabulary will be integrated in the Metadata Editor developed for the Reference Data and Service Initiative of JRC Institute for Environment and Sustainability (JRC-IES RDSI). Additionally, the importance of maintaining the Drought Vocabulary alive has been highlighted. In the last internal meeting of WP5, it was agreed that EDO should be responsible for the maintenance and update of the Drought Vocabulary beyond the scope of the EuroGEOSS project. A contact point will be established to inform interested third parties on updates of the vocabulary. Details and modus operandi will be discussed in the next months.

REFERENCES

American Meteorological Society. (2000). *Glossary of Meteorology, Second Edition*. Todd S. Glickman (Ed). Boston, United States: AMS Books.

ANSI/NISO. (2005). *Guidelines for the Construction, Format, and Management of Monolingual Thesauri.* American National Standards Institute (ANSI), Z39-19.

Bechhofer, S. and C. Goble. (2001). Thesaurus construction through knowledge representation. *Data & Knowledge engineering*, 37: 25-45.

BSI. (2007). *Structured vocabularies for information retrieval. Guide*. British Standards Institute (BSI), BS 8723.

Commonwealth of Australia. (2001). The DIRKS methodology: A users guide, at: http://www.naa.gov.au/Images/dirks_part1_tcm2-935.pdf [accessed 17 August 2011].

Commonwealth of Australia. (2003). *Developing a Functions Thesaurus. Guidelines for Commonwealth Agencies*, at: http://www.naa.gov.au/images/developing-a-thesaurus_tcm2-916.pdf [accessed 16 August 2011].

EuroGEOSS. (2010). *Deliverable Description of the Initial Operating Capacity*, at: http://www.eurogeoss.eu/wp/wp5/Documents/D.5.5_InitialOperatingCapacity_DeliverableDescription_FINAL.pdf [accessed 12 August 11].

European Commission. (2007). Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 Establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). *Official Journal of the European Union*, 50 (L 108) of 25 April 2007: 1-14.

European Commission. (2008). Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata *Official Journal of the European Union,* (L 326), of 4 December 2008: 12–30.

GEO, Group of Earth Observation. (2009). *GCI Consolidated Requirements*, at: http://www.earthobservations.org/documents/gci/gci_requirements_20090312.doc [accessed 9 August 2011].

Gruber, T. R. (1993). A translation approach to portable ontology specifications, *Knowledge Acquisition*, 5(2): 199–220.

ISO. (1986). Documentation: Guidelines for the establishment and development of monolingual thesauri. ISO 2788, International Organization for Standardization.

ISO. (2011). *Thesauri and interoperability with other vocabularies (draft)*. International Organization for Standarization (ISO), ISO 25964.

Lacasta, J., J. Nogueras-Iso, F. J. Zarazaga-Soria. (2010) *Terminological Ontologies: Design, Management and Practical Applications*. Germany: Springer.

Lacasta, J., J. Nogueras-Iso, R. Béjar, P. R. Muro-Medrano and F. J. Zarazaga-Soria. (2007). A Web Ontology Service to facilitate interoperability within a Spatial Data Infrastructure: applicability to discovery*, Data & Knowledge Engineering*, 63: 945-969.

Latre, M. Á., Lacasta, J., Mojica, E., Nogueras-Iso, J. and Zarazaga-Soria, F. J. (2009). "An Approach to Facilitate the Integration of Hydrological Data by means of Ontologies and Multilingual Thesauri" in Sester M., Bernard L. and Paelke V. (Eds). *Advances in GIScience.* Springer Berlin Heidelberg, pp. 155–171.

Latre, M. Á., Nogueras-Iso, J. and Hofer, B. (2011a). "Searching drought-related resources through a specialized vocabulary - testing the interoperable infrastructure of the EuroGEOSS project". *Proceedings of the INSPIRE Conference 2011: INSPIREd by 2020 - Contributing to smart, sustainable and inclusive growth*. June 27 - July 1. Edinburgh, Scotland.

Latre, M. Á., Hofer, B., Lacasta, J. and Nogueras-Iso, J. (2011b). The development and interlinkage of a drought vocabulary in the EuroGEOSS interoperable infrastructure. Currently under review for the *International Journal of Spatial Data Infrastructures Research*, at: http://ijsdir.jrc.ec.europa.eu/index.php/ijsdir/article/view/264 [accessed 11 November 2011].

Miles, A. and Pérez-Agüera, J.R. (2007). SKOS: Simple Knowledge Organisation for the Web. *Cataloging and Classification Quarterly*, 43(3–4): 69–83.

Nogueras-Iso, J., Barrera, J., Gracia-Crespo, F., Laiglesia, S. and Muro-Medrano, P. R. (2008). "Integrating catalog and GIS tools: access to resources from CatMDEdit thanks to gvSIG", *Proceedings of the 4th International gvSIG conference: moving forward together.* December 3-5 2008, Valencia, Spain.

Semantic Health Project. (2006). *The creation of the Belgian Bilingual Bi-encoded Thesaurus (3BT).* World Health Organization. at: http://www.semantichealth.org/PUBLIC/Belgium_The%20creation%20of%203BT.pdf [accessed 16 August 2011].

Sowa, J. F. (1996). Ontologies for Knowledge Sharing. In *Manuscript of the invited talk at Terminology and Knowledge Engineering Congress (TKE '96),* Vienna.

State Records Authority of New South Wales. (2003). *Guidelines for Developing and Implementing a Keyword Thesaurus*, at: http://www.records.nsw.gov.au/recordkeeping/government-recordkeeping-manual/documents/recordkeeping-guidelines/Developing%20and%20Implementing%20a%20Keyword%20Thesaurus.pdf [accessed 16 August 2011].

UNESCO. (1993). *International Glossary of Hydrology / Glossaire International D'hydrologie* (Wmo/Omm/Vmo, No. 385), United States: Unipub.

Uschold, M. and M. Gruninger. (2004). Ontologies and semantics for seamless connectivity, *ACM SIGMOD Record*, 33(4): 58–64.

Vickery, B. C. (1966). *Faceted Classification Schemes*. Rutges Series on Systems for the Intellectual Organization on Information, Rutgers State University, New Brunswick, NK,

Vorsey, K. L. De, C. Elson, N. P. Gregorev and J. Hansen. (2006). The development of a local thesaurus to improve access to the anthropological collections of the American Museum of Natural History. *D-Lib Magazine*, 12 (4).

Working Group on Guidelines for Multilingual Thesauri of the IFLA Classification and Indexing Section. (2005). *Guidelines for Multilingual Thesauri*, IFLA Professional Reports, No. 115.

International Federation of Library Associations and Institutions, at: http://archive.ifla.org/VII/s29/pubs/Profrep115.pdf [accessed 16 August 2011].